Econometric Analysis

Ch.4 The Single-Equation Linear Model and OLS Estimation

Ryuichi Tanaka

4.1 Overview of the Single-Equation Linear Model

The population model : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + u$

$(y, \mathbf{x})$: observable random vector

u : unobservable random disturbance

$\boldsymbol{\beta}$: parameters to be estimated

Assumptions for the consistency of OLS estimator

(1) $E(u) = 0$ (automatically satisfied if the model has a constant term)

(2) $Cov(x_j, u) = 0,\ j = 1, 2, \cdots, K.$ (no correlation between u and x)

Endogeneity

Definition

In the population model, x_j is said to be endogenous (exogenous) if $Cov(x_j, u) \neq 0$ ($Cov(x_j, u) = 0$).

- Three main causes of endogeneity
  - Omitted variables
  - Measurement error
  - Simultaneity

Omitted variable

Popuolation model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_q q + \varepsilon$

Estimable model: $y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + u$ where $u = \beta_q q + \varepsilon$

q is unobservable.

If q is correlated with x, x is endogenous

(e.g., x_1 : education, q: ability)

(See section 4.3)

Measurement error

Popuolation model: $y = \beta_0 + \beta x^* + \varepsilon$

Estimable model: $y = \hat{\beta}_0 + \hat{\beta} x + u$ where $u = \varepsilon - \tilde{\varepsilon}$ and $x = x^* + \tilde{\varepsilon}$

x^* is observed with measurement error

If $\mathrm{cov}(x, \tilde{\varepsilon}) \neq 0$, then $Cov(x, u) \neq 0$

(e.g., x^*:income, y:saving)

See section 4.4

Simulataneity

The population model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,\ x_1 = \delta_0 + \delta_1 y$

y : murder rate, x_1 : number of police officers

If you estimate the model

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$

ignoring $x_1 = \delta_0 + \delta_1 y$, then

$Cov(x_1, u) = Cov(\delta_0 + \delta_1 y, u) = Cov(\delta_0 + \delta_1(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + u), u) \neq 0$

See chapter 9

4.2 Asymptotic Properties of OLS

- Consistency?
- Asymptotic distribution?

The population model : $y = \mathbf{x}\boldsymbol{\beta} + u$, $\mathbf{x}$: 1 x K with $x_1 = 1$, $\boldsymbol{\beta}$: K x 1

$\{(\mathbf{x}_i, y_i) : i = 1, 2, \cdots, N\}$: a random sample of size N from the population (to estimate beta), each observatio n is i.i.d.

For each $(\mathbf{x}_i, y_i)$, we have $y_i = \mathbf{x}_i\boldsymbol{\beta} + u_i$

4.2.1 Consistency

Assumption OLS.1 (population orthogonality condition): $E(\mathbf{x}'u) = \mathbf{0}$

This assumption is equivalent to $E(u) = 0$ (with $x_1 = 1$) and $Cov(x_j, u) = 0$

Mean independence ($E(u \mid \mathbf{x}) = 0$) is stronger than **AOLS.1**

($E(u \mid \mathbf{x}) = 0 \Rightarrow$ **AOLS.1**)

AssumptionOLS.2 (full rank condition): $rank E(\mathbf{x}'\mathbf{x}) = K$.

There is no perfect colinearity among regressors in the population

Identifiability

β is identifiable if β can be written in terms of population moments in observable variables

Claim: Under AOLS.1 and AOLS.2, β is identifiable

$$\mathbf{x}'y = \mathbf{x}'\mathbf{x}\boldsymbol{\beta} + \mathbf{x}'u$$
$$E(\mathbf{x}'y) = E(\mathbf{x}'\mathbf{x}\boldsymbol{\beta}) + E(\mathbf{x}'u) = E(\mathbf{x}'\mathbf{x})\boldsymbol{\beta}$$
$$\boldsymbol{\beta} = [E(\mathbf{x}'\mathbf{x})]^{-1} E(\mathbf{x}'y)$$

AOLS.1 is needed for consistency

AOLS.2 is needed for identification

The analogy principle

Turn the popyulation problem into its sample counterpart

Replace $E(\mathrm{x}'\mathrm{x})$, $E(\mathrm{x}'y)$ by $N^{-1}\sum_{i=1}^{N}\mathrm{x}_i'\mathrm{x}_i$, $N^{-1}\sum_{i=1}^{N}\mathrm{x}_i'y_i$

(1) $\hat{\beta} = [N^{-1}\sum_{i=1}^{N}\mathrm{x}_i'\mathrm{x}_i]^{-1}[N^{-1}\sum_{i=1}^{N}\mathrm{x}_i'y_i] = \beta + [N^{-1}\sum_{i=1}^{N}\mathrm{x}_i'\mathrm{x}_i]^{-1}[N^{-1}\sum_{i=1}^{N}\mathrm{x}_i'u_i]$

(2) Under AOLS2, $p\,lim([N^{-1}\sum_{i=1}^{N}\mathrm{x}_i'\mathrm{x}_i]^{-1}) = \mathrm{A}^{-1}$ where $\mathrm{A} \equiv E(\mathrm{x'x})$

(3) Under AOLS1, $p\,lim(N^{-1}\sum_{i=1}^{N}\mathrm{x}_i'u_i) = E(\mathrm{x}'u) = 0$

(4) With Slutsky's theorem, $p\,lim(\hat{\beta}) = \beta + \mathrm{A}^{-1}0 = \beta$

Theorem 4.1: Consistency of the OLS

Theorem 4.1 (consistency of OLS):
Under AOLS.1 and AOLS.2, the OLS estimator $\hat{\boldsymbol{\beta}}$ obtained from a random sample following the population model is consistent for $\boldsymbol{\beta}$.

4.2.2: Asymptotic Inference Using OLS

$$\sqrt{N}(\hat{\beta}-\beta) = \left[N^{-1}\sum_{i=1}^{N}\mathrm{x}_i'\mathrm{x}_i\right]^{-1}\left[N^{-1/2}\sum_{i=1}^{N}\mathrm{x}_i'u_i\right]$$

(1) WLLN implies $p\,lim\left[N^{-1}\sum_{i=1}^{N}\mathrm{x}_i'\mathrm{x}_i\right]^{-1} = \mathrm{A}^{-1}$

(2) Since $\{(\mathrm{x}_i'u_i): i = 1,2,\cdots\}$ are *i.i.d.* sequence with zero mean,
CLT implies that $\left[N^{-1/2}\sum_{i=1}^{N}\mathrm{x}_i'u_i\right] \xrightarrow{d} Normal(0,\mathrm{B})$

(3) Hence $\sqrt{N}(\hat{\beta}-\beta) \xrightarrow{d} Normal(0,\mathrm{A}^{-1}\mathrm{B}\mathrm{A}^{-1})$ where $\mathrm{B} \equiv E(u^2\mathrm{x'x})$

$$A\,\hat{\mathrm{var}}(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}/N$$

Homoskedasticity

Assumption OLS.3 (homoskedasticity):

$E(u^2 \mathbf{x}'\mathbf{x}) = \sigma^2 E(\mathbf{x}'\mathbf{x})$ where $\sigma^2 \equiv E(u^2)$

Theorem 4.2 (Asymptotic Normality of OLS) :

Under AOLS.1, 2, and 3, $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} Normal(0, \sigma^2 \mathbf{A}^{-1})$

Under AOLS3, asymptotic variance has the simple form.

Plugging consistent estimators of σ^2 and A,

$A\hat{\mathrm{var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$ [where $\mathbf{X}'\mathbf{X} = \sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i$]

This is the "usual" variance-covariance matrix of OLS estimator

t-test and F-test are valid in the limit (asymptotically valid)

4.2.3 Heteroskedasticity-Robust Inference

- Under AOLS.1 and 2, OLS estimator is consistent (even without AOLS3)
- Without AOLS.3, the “usual” OLS variance-covariance matrix (and thus the standard errors) are not asymptotically valid (hence we cannot use t-test or F-test)
- Solution : use heteroskedasticity-robust variance matrix (standard error, test statistics)

Robust variance matrix

Without AOLS3, $Avar(\hat{\beta}) = \mathrm{A}^{-1}\mathrm{B}\mathrm{A}^{-1}/N$, $\mathrm{A} \equiv E(\mathbf{x}'\mathbf{x})$, $\mathrm{B} \equiv E(u^2\mathbf{x}'\mathbf{x})$

Estimator of asymptotic variance: $\hat{\mathrm{V}} = A\hat{var}(\hat{\beta}) = \hat{\mathrm{A}}^{-1}\hat{\mathrm{B}}\hat{\mathrm{A}}^{-1}/N$

Consistent estimator of $\mathrm{A} \equiv E(\mathbf{x}'\mathbf{x})$: $\hat{\mathrm{A}} = \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i \right]$

Consistent estimator of $\mathrm{B} \equiv E(u^2\mathbf{x}'\mathbf{x})$:

$\hat{\mathrm{B}} = N^{-1} \sum_{i=1}^{N} \hat{u}_i^2 \mathbf{x}_i'\mathbf{x}_i$ where $\hat{u}_i = y_i - \mathbf{x}_i\hat{\beta}$ (OLS residuals):

Test of linear constraint $H_0 : \mathrm{R}\beta = \mathrm{r}$:

$W = (\mathrm{R}\hat{\beta} - \mathrm{r})'(\mathrm{R}\hat{\mathrm{V}}\mathrm{R}')^{-1}(\mathrm{R}\hat{\beta} - \mathrm{r}) \sim \chi^2_Q$

4.2.4 Lagrange Multiplier (Score) Test

The population model : $y = \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + u$

Null hypothesis : $H_0 : \beta_2 = 0$

LM Test

1. Estimate $y = \mathbf{x}_1\beta_1 + u$ by OLS and obtain OLS residual $\tilde{u}_i = y_i - \mathbf{x}_{i1}\tilde{\beta}_1$
2. Regress $\tilde{u}_i$ on $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$, and obtain the usual R_u^2. Then

$LM = NR_u^2 \sim \chi^2_{K_2}$ (K_2 is the number of restrictions)

Interpretation

If the null hypothesis $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ is correct, $\tilde{u}_i$ and $\mathbf{x}_{2i}$ are uncorrelated

($\mathbf{x}_{2i}\boldsymbol{\beta}$ is not in the error term)

If $\tilde{u}_i$ and $\mathbf{x}_{2i}$ are correlated, $\mathbf{x}_{2i}$ explains $\tilde{u}_i$ well and thus R_u^2 becomes large

(and thus reject the null hypothesis)

LM Test

LM Test under homoskedasticity

$$LM = \left(N^{-1/2}\sum_{i=1}^{N}\hat{r}_i'\tilde{u}_i\right)'\left(\tilde{\sigma}^2 N^{-1}\sum_{i=1}^{N}\hat{r}_i'\hat{r}_i\right)^{-1}\left(N^{-1/2}\sum_{i=1}^{N}\hat{r}_i'\tilde{u}_i\right)$$

$\tilde{\sigma}^2 = N^{-1}\sum_{i=1}^{N}\tilde{u}_i^2$, $\hat{r}_i$: 1 x K vector of OLS residuals from the regression of x_{2i} on x_{1i}

LM test under heteroskedasticity

$$LM = \left(N^{-1/2}\sum_{i=1}^{N}\hat{r}_i'\tilde{u}_i\right)'\left(N^{-1}\sum_{i=1}^{N}\tilde{u}_i^2\hat{r}_i'\hat{r}_i\right)^{-1}\left(N^{-1/2}\sum_{i=1}^{N}\hat{r}_i'\tilde{u}_i\right)$$

How to obtain the test statistics

1. Regress x_{2i} on x_{1i} and collect the residuals in $\hat{r}_i$
2. Form $\tilde{u}_i \cdot \hat{r}_i$ and run the regression 1 on $\tilde{u}_i \cdot \hat{r}_i$ and obtain SSR_0.

Then $N - SSR_0 \sim \chi^2_{K_2}$

4.3 OLS Solutions to the Omitted Variables Problem
4.3.1 OLS Ignoring the Omitted Variables

The error form of the structural model with omitted variable q :

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \gamma q + v,\ E(v \mid x_1, x_2, \cdots, x_K, q) = 0$

Example y: log(wage), q: ability, x: education

Estimable model:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + u,\quad u = \gamma q + v$

If q and x are correlated, u and x are correlated (endogeneity)

OLS estimator is inconsistent

OLS Omitted Variable Inconsistency

Linear projection of q on observables (1 and $x_1, \cdots x_K$)

$q = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_K x_K + r$

$E(r) = 0,\ \ Cov(x_i, r) = 0,\ \ \boldsymbol{\delta} = [Var(\mathbf{x})]^{-1} Cov(\mathbf{x}, q),\ \ \delta_0 = q - E(\mathbf{x})\boldsymbol{\delta}$

Estimable equation :

$y = (\beta_0 + \gamma\delta_0) + (\beta_1 + \gamma\delta_1)x_1 + (\beta_2 + \gamma\delta_2)x_2 + \cdots + (\beta_K + \gamma\delta_K)x_K + u,\ \ u = \gamma r + v$

$E(u) = 0,\ \ Cov(x_i, u) = 0 \Rightarrow p\lim(\hat{\beta}_j) = \beta_j + \gamma\delta_j$

Suppose that only x_K is endogenous

$p\lim(\hat{\beta}_j) = \beta_j$ for $j = 1, \ldots, K-1$

$p\lim(\hat{\beta}_K) = \beta_K + \gamma\delta_K = \beta_K + \gamma[Var(x_K)]^{-1} Cov(x_K, q)$

We can infer the sign and magnitude of asymptotic bias from this equation

Example 4.2 Over-estimation of the returns to education

Structural model in the error form:

$\log(wage) = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 educ + \gamma abil + v$

abil is unobservable

LP of abil: $abil = \delta_0 + \delta_1 exper + \delta_2 exper^2 + \delta_3 educ + r$

Under $\delta_1 = \delta_2 = 0$, β_1 and β_2 can be consistently estimated

However, $p\lim \hat{\beta}_3 = \beta_3 + \gamma\delta_3$

The returns to education is overestimated

if ability raises wages （γ>0) and education is positively

correlated with ability （δ_3>0)

4.3.2 The Proxy Variable-OLS Solution

If you have proxy variables to unobservables (omitted variables), you can eliminate (or mitigate at least) omitted variable bias

Proxy variable

A variable proxies for unobserved variables, but not contained in the population model (structure equation)

Ex IQ test score is a proxy variable for ability

Two formal requirements for a Proxy variable

(Requirement 1) Proxy variable (z) is redundant (ignorable) in the structural equation:

$$E(y \mid \mathbf{x}, q, z) = E(y \mid \mathbf{x}, q)$$

After controlling x and q, z has no explanation power for y

Ex. z : IQ test score

Since not IQ test score but ability does increase wage, IQ test score has no explanation power for wage after controlling ability

Two formal requirements for a Proxy variable

(Requirement 2) Correlation between the omitted variable q and each xj be zero once we partial out z. Using LP,

$$L(q \mid 1, x_1, \cdots, x_K, z) = L(q \mid 1, z)$$

Z is closely enough related to unobserved variable q so that once it is included in the above LP, the other xs' are not partially correlated with q

Proxy variable is a variable that proxies for unobserved variables but not contained in the structural equation

(Perfect) proxy variable can eliminate omitted variable bias

LP: $q = \theta_0 + \theta_1 z + r$, $E(r) = 0$, $Cov(z, r) = 0$ (from the definition of LP)

If z is a proxy variable, $\theta_1 \neq 0$ and $Cov(x_j, r) = 0$

Then estimable equation is $y = (\beta_0 + \gamma\theta_0) + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \gamma\theta_1 z + (\gamma r + v)$

$E(v \mid x_1, x_2, \cdots, x_K, q) = 0$ and $Cov(x_j, r) = 0$ imply $Cov(x_j, u) = 0$

The requirement 1 implies $Cov(z, v) = 0$

(without requirement 1, z and v are correlated because z is contained in v as an omitted variable)

Since $Cov(z, r) = 0$ by definition, $Cov(z, u) = 0$

Therefore, the OLS estimator of $y = (\beta_0 + \gamma\theta_0) + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \gamma\theta_1 z + u$ is *consistent*!

Imperfect proxy variable

If some of x_j correlates to r in $q=\theta_0+\theta_1 z+r$ (by the failure of the requiremt 2),

$q=\theta_0+\rho_1 x_1+\cdots+\rho_K x_K+\theta_1 z+r$

Then, $p\lim(\hat{\beta}_j)=\beta_j+\gamma\rho_j$となる

In this case, the OLS estimator is inconsistent, but asymptotic bias is small as long as z is a good proxy variable (not perfect though) with small ρ_j

4.4 Properties of OLS under Measurement Error

4.4.1 Measurement Error in the Dependent Variable

Measurement error in explained variable

$y^*=\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_K x_K+v$

y^*: true value, $y\neq y^*$: observed value, $e_0=y-y^*$: population measurement error

Estimable model

$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_K x_K+v+e_0$

OLS estimators are consistent if the measurement error is uncorrelated to the regressors

Only cost is larger variance of the error term

4.4.2 Measurement Error in the Explanatory Variable

$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_K x_K^*+v$

x_K^*: true value, $x_K\neq x_K^*$: observed value

Assume $x_K\neq x_K^*$ is uncorrelated to v （redundancy assumption）

$e_K=x_K-x_K^*$:population measurement error with zero mean

$E(x_j e_K)=0$ for j=1,…,K-1

<u>Case 1</u> $Cov(x_K,e_K)=0$

$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_K x_K+(v-\beta_K e_K)$

The error term is uncorrelated to regressors

OLS estimator is consistent

Classical Errors in variable (CEV)

Case 2 $Cov(x_K^*,e_K)=0$

$Cov(x_K,e_K)=E(x_K e_K)=\sigma_{e_K}^2$

$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_K x_K+(v-\beta_K e_K)$ OLS estimator is inconsistent

$p\lim(\hat{\beta}_K)=\beta_K\left(\frac{\sigma_{r_K^*}^2}{\sigma_{r_K^*}^2+\sigma_{e_K}^2}\right)$ where r_K^* is the linear projection error in

$x_K^*=\delta_0+\delta_1 x_1+\delta_2 x_2+\cdots+\delta_{K-1}x_{K-1}+r_K^*$

With CEV, OLS estimator "shrinks" toward zero (attenuation bias)

In the case with only one regressor,

$p\lim(\hat{\beta}_1)=\beta_1\left(\frac{\sigma_{x_1^*}^2}{\sigma_{x_K^*}^2+\sigma_{e_1}^2}\right)=\beta_1\left(\frac{Var(x_1^*)}{Var(x_1)}\right)$

PS3 and PS4

PS3

- **4.1**
- **4.6**
- **4.7**
- **Reproduce the estimation results in Example 4.1 with Stata (hand in both the output and the DO files)**

PS4

- **Reproduce the estimation results in Example 4.3 and Example 4.4 with Stata (hand in both the output and the DO files)**
- **4.9**
- **4.11**
- **4.14**